Validity of composite end points in clinical trials

Victor M Montori, Gaietà Permanyer-Miralda, Ignacio Ferreira-González, Jason W Busse, Valeria Pacheco-Huergo, Dianne Bryant, Jordi Alonso, Elie A Akl, Antònia Domingo-Salvany, Edward Mills, Ping Wu, Holger J Schünemann, Roman Jaeschke, Gordon H Guyatt

Use of composite end points as the main outcome in randomised trials can hide wide differences in the individual measures. How should you apply the results to clinical practice?

Department of Medicine, Mayo Clinic College of Medicine, Rochester, MN 55905, USA
Victor M Montori *assistant professor*

Cardiology Service, Epidemiology Unit, Hospital General Vall d'Hebron, Barcelona 08035, Spain
Gaietà Permanyer-Miralda *senior consultant*

Department de Medicina, Universitat Autònoma de Barcelona, Hospital General Vall d'Hebron
Ignacio Ferreira-González *research fellow*

Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, ON L8N 3Z5, Canada
Jason W Busse *research associate*
Dianne Bryant *instructor*
Gordon H Guyatt *professor*

Department of Medicine, McMaster University
Roman Jaeschke *clinical professor*

Primary Care Centre Turó, EAP Vilapicina, Institut Català de la Salut, Barcelona, Spain
Valeria Pacheco-Huergo *general practitioner*

continued over

BMJ 2005;330:594–6

Improvements in medical care over the past two decades have decreased the frequency with which patients with common conditions such as myocardial infarction develop subsequent adverse events. Although welcome for patients, low event rates provide challenges for clinical investigators, who consequently require large sample sizes and long follow up to test the incremental benefits of new treatments. Clinical trialists have responded to these challenges by relying increasingly on composite end points, which capture the number of patients experiencing any one of several adverse events—for example, death, myocardial infarction, or hospital admission.[1]

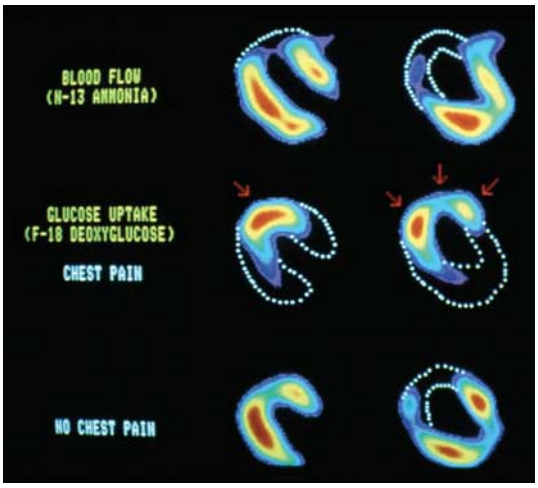
Use of composite end points is usually justified by the assumption that the effect on each of the components will be similar and that patients will attach similar importance to each component.[1] But this is not always the case. In this article we provide a strategy to interpret the results of clinical trials when investigators measure the effect of treatment on an aggregate of end points of varying importance.

Example case

Consider a 76 year old man who has disabling angina despite taking β blockers, nitrates, aspirin, an angiotensin converting enzyme inhibitor, and a statin. His doctor suggests cardiac catheterisation and possible revascularisation. The patient is reluctant to have invasive management, and wonders how much benefit he might expect from surgery.

The trial of invasive versus medical therapy in elderly patients (TIME) is relevant.[2] The study randomised 301 patients aged 75 years or older with resistant angina to optimised drug treatment or cardiac catheterisation and possible revascularisation. Although the groups showed no difference in quality of life at 12 months, the frequency of a composite end point (death, non-fatal myocardial infarction, and hospital admission for acute coronary syndrome) was much lower in the revascularisation group (25.5%) than in the medical management arm (64.2%; hazard ratio 0.31, 95% confidence interval 0.21 to 0.45).

Although the overall result suggests invasive treatment would be beneficial, marked differences existed in the absolute reduction in risk across components (table 1). In the invasive group, five more patients died but there were six fewer myocardial infarctions and 78 fewer hospital admissions. How should you interpret these results and inform the patient?



Scan of heart of patient with angina: much of treatment evidence relies on composite end points

Evaluating composite end points

Clinicians can use three questions to help decide whether to base a clinical decision on the effect of treatment on a composite end point or on the component end points (box). We will not expand on statistical issues here, but box A on bmj.com gives a brief outline.

Importance of individual components to patients

When all components of a composite end point are of equal importance to the patient, it will not be misleading to assume that the effect of the intervention on each component is similar, in both relative and absolute terms. If patients consider death, stroke, and myocardial infarction of equal importance, it does not much matter how a 5% absolute risk reduction in the

Further examples are available on bmj.com

Guide to interpreting composite end points

1. Are the component end points of similar importance to patients?
2. Did the more and less important end points occur with similar frequency?
3. Are the component end points likely to have similar relative risk reductions?
   Is the underlying biology of the component end points similar?
   Are the point estimates of the relative risk reductions similar and the confidence intervals sufficiently narrow?

The extent to which the answers to these questions are no will determine whether you need to examine the component end points separately

composite end point is distributed. The decision will be the same, even if treatment effects differ substantially.

Patients almost invariably, however, assign varying importance to different health outcomes. As a result, we can seldom ignore possible differences in treatment effects between component end points on the grounds that patients give them identical importance. The magnitude of the gradient in importance between end points therefore becomes the issue.

For instance, consider a trial of four doses of perioperative aspirin in patients having carotid endarterectomy in which the composite end point included death and stroke.[3] Many patients would consider a stroke as having a negative value approaching that of death. The relatively small gradient in importance between the components increases the likely usefulness of the composite end point in clinical decision making. In a trial of corticosteroids among patients with acute exacerbation of chronic obstructive lung disease, however, the investigators chose a combined end point of death from any cause, need for intubation and mechanical ventilation, and administration of unblinded steroids.[4] Patients are likely to consider the need for short term steroids of trivial importance compared with mechanical ventilation and death, raising questions about the suitability of combining these components.

Frequency of component end points

The heart outcomes prevention evaluation (HOPE) study randomised 9297 patients at high risk of cardiac events to ramipril or placebo.[5] Ramipril reduced cardiovascular deaths from 8.1% to 6.1% (relative risk reduction 26%, 95% confidence interval 13% to 36%), myocardial infarction from 12.3% to 9.9% (20%, 10% to 30%), and stroke from 4.9% to 3.4% (32%, 16% to 44%). The gradient in rates of death, myocardial infarction, and stroke in the control group (8.1%, 12.3%, and 4.9%) is relatively small. The difference in events between treatment and control (2.0% for deaths, 2.4% for myocardial infarction, and 1.5% for stroke) is even more similar. This provides support for focusing on the composite end point in clinical decision making. In some studies, however, the frequency of component end points differs greatly (see box B on bmj.com).

Treatment effects on component end points

Confidence in a composite end point rests partly on a belief that similar reductions in relative risk apply to all the components. Investigators should therefore construct composite end points in which the biology would lead us to expect similar effects across components.

For example, the authors of the CAPRIE study, a randomised trial of aspirin versus clopidogrel in patients at risk of ischaemic events, argued explicitly for the biological sense of their composite end point.[6] Citing results of a meta-analysis of 142 trials of antiplatelet drugs versus placebo, they note the similar biological determinants of ischaemic stroke, myocardial infarction, and vascular death. Their argument strengthens the case for assuming, barring contrary evidence, that relative risk reductions are consistent across components of the composite end point. Box C on bmj.com describes a trial in which biology argues against expecting similar relative risk reductions across component end points.

No matter how strong the biological rationale, only evidence of similar relative risk reductions can strongly increase our comfort with a composite end point. In the HOPE trial described above, the risk reductions were similar for all components.[5] In the losartan intervention for end point reduction in hypertension study, however, the relative risk reductions of the component end points differed greatly (−7% for myocardial infarction, 25% for stroke, and 11% for cardiovascular death),[7] even though the rationale for using a combined end point was the same as in the CAPRIE study.[6] In this case it would be better to consider individual component end points. Sometimes the risk reductions of component end points look similar but the confidence intervals are wide (box D on bmj.com).

Applying the questions

Let us return to the scenario of the patient reluctant to have surgery to control his angina. Is it reasonable to use the composite end point from the TIME trial (death, myocardial infarction, and hospital admission for acute coronary syndrome) to guide the decision, or should we focus on individual results of the three components?

To determine the answer, we can ask the three questions in the box. In response to the first question, most patients will find death and serious myocardial

Health Services Research Unit, Institut Municipal d'Investigació Mèdica, Barcelona, Spain
Jordi Alonso *head*
Antònia Domingo-Salvany *senior scientist*
Department of Medicine and of Social and Preventive Medicine, University at Buffalo, Buffalo, NY 14214, USA
Elie A Akl *research assistant professor*
Holger J Schünemann *associate professor*
Department of Research, Canadian College of Naturopathic Medicine, Toronto, Canada M2K 1E2
Ping Wu *research assistant*
Edward Mills *director of research*
Correspondence to: G Permanyer-Miralda gpermany@vhebron.net

Table 1 Results from the TIME trial of invasive versus medical treatment for angina[1]

End point	Invasive (n=153)	Medical (n=148)	% Risk difference (95% CI)	Hazard ratio (95% CI)
Composite end point*	39 (25.5%)	95 (64.2%)	38.7 (27.9% to 48.5%)	0.31 (0.21 to 0.45)
Death	17	12	−3.0% (−9.9% to 3.8%)	1.51 (0.72 to 3.16)
Non-fatal myocardial infarction†	14	20		0.75 (0.36 to 1.55)
Admission for acute coronary syndrome†	28	106		0.19 (0.12 to 0.30)

*Mortality, non-fatal myocardial infarction, and hospital admission for acute coronary syndrome.
†The authors report number of events rather than patients.

Summary points

Composite end points are outcomes that capture the number of patients experiencing one or more of several adverse events

The validity of composite end points depends on similarity in patient importance, treatment effect, and number of events across the components

When large variations exist between components the composite end point should be abandoned

infarction with subsequent disability far more important than a short admission for acute coronary syndrome with rapid return to previous function.

The answers to the other two questions are also negative. Hospital admissions occurred far more frequently than the two more important events (table). Biological rationale fails to support a presumption that the invasive strategy will have similar effects on all three end points. Indeed, the investigators explicitly state that they expect an increase in short term deaths with surgery, while achieving benefits in terms of decreased angina and associated hospital admissions. The trend toward increased deaths, with a large reduction in admissions, with the invasive strategy provides support for this hypothesis. The composite end point thus fails all three criteria and provides little useful information for clinical decision making.

Conclusions

The widespread use of composite end points reflects their elegant simplicity as a solution to the problem of declining event rates. Unfortunately, use of composite end points makes the interpretation of the results of randomised trials for clinical decision making challenging. Investigators and their sponsors may claim treatment effects over a broad range of outcomes, whereas the effect may in fact be limited to one component. Occasionally, composite end points prove useful and informative for clinical decision making. Often, they do not. These users' guides will help clinicians differentiate between these situations.

Contributors and sources: The authors are clinicians, methodologists, and trialists with expertise in the conduct or the interpretation of clinical trials. In preparation for this article we reviewed Medline and the Cochrane Methodology Register for studies, editorials, and commentaries about the use of composite end points in clinical trials. GP-M, IF-G, and GHG conceived the idea for the article; VMM, GP-M, IF-G, and GHG reviewed the methodological literature and contributed to the framework presented in this manuscript; VMM and GHG created the first draft, and edited subsequent revisions; all authors offered critical revisions to the manuscript and the illustrative examples we used in the manuscript; all approved the final version; GHG is the guarantor.

Funding: VMM is a Mayo Foundation Scholar. JWB is funded by a Canadian Institutes of Health research fellowship award. G P-M and JA are partially supported by funds from the Fondo de Investigación Sanitaria. EM is funded by an Ontario HIV Treatment Network research award.

Competing interests: None declared.

1 Freemantle N, Calvert M, Wood J, Eastaugh J, Griffin C. Composite outcomes in randomized trials: greater precision but with greater uncertainty? *JAMA* 2003;289:2554-9.
2 TIME Investigators. Trial of invasive versus medical therapy in elderly patients with chronic symptomatic coronary-artery disease (TIME): a randomised trial. *Lancet* 2001;358:951-7.
3 Taylor DW, Barnett HJ, Haynes RB, Ferguson GG, Sackett DL, Thorpe KE, et al. Low-dose and high-dose acetylsalicylic acid for patients undergoing carotid endarterectomy: a randomised controlled trial. ASA and Carotid Endarterectomy (ACE) Trial Collaborators. *Lancet* 1999;353:2179-84.
4 Niewoehner D, Erbland M, Deupree R, Collins D, Gross N, Light R, et al. Effect of systemic glucocorticoids on exacerbations of chronic obstructive pulmonary disease. *N Engl J Med* 1999;340:1941-7.
5 Yusuf S, Sleight P, Pogue J, Bosch J, Davies R, Dagenais G. Effects of an angiotensin-converting-enzyme inhibitor, ramipril, on cardiovascular events in high-risk patients. The Heart Outcomes Prevention Evaluation Study Investigators. *N Engl J Med* 2000;342:145-53.
6 CAPRIE Steering Committee. A randomised, blinded, trial of clopidogrel versus aspirin in patients at risk of ischaemic events (CAPRIE). *Lancet* 1996;348:1329-39.
7 Dahlof B, Devereux RB, Kjeldsen SE, Julius S, Beevers G, Faire U, et al. Cardiovascular morbidity and mortality in the losartan intervention for endpoint reduction in hypertension study (LIFE): a randomised trial against atenolol. *Lancet* 2002;359:995-1003.

(Accepted 10 January 2005)

Corrections and clarifications

Population based randomised controlled trial on impact of screening on mortality from abdominal aortic aneurysm
Tables 2 and 3 of both the abridged and the full versions of this paper by Paul E Norman and colleagues contain some incorrect values (*BMJ* 2004;329:1259-62). In table 2, for the emergency procedures the "all ruptures" values are 19 and 22 for the "not scanned" and "total invited" groups respectively and 27 in the control group; in table 3, the corresponding values are 32, 35, and 38. The authors state, however, that this amendment does not alter their analyses or conclusions.

Risk of ischaemic stroke in people with migraine: systematic review and meta-analysis of observational studies
In our haste to correct this paper by Etminan and colleagues (*BMJ* 2005;330:63-5, 8 Jan), some of the authors' late corrections were not carried out properly—either at proof stage of the abridged print version or in the correction that we subsequently published on the web relating to the full version only. In the abridged version, the relative risk for migraine with aura in table 2 should be 2.88 (not 2.28); in table 1, the upper confidence limit for migraine with aura for Schwaag should be 3.53 (not 3.35), and the cases:controls for the Collaborative Group should be 430:151 (not 430:451). In the results section of the full version, the references for the data on migraine with and without aura are numbers 2, 3, 12-14, 17-19; in table 1, the cases:controls with migraine is 26:26 for Donaghy (not for Chang as stated in the previous correction).

Minerva
The eighth Minerva item (about a study published in *Neurology*) in the issue of 22 January (*BMJ* 2005;330:204) may have misled readers by including as its first sentence: "Survival in patients with Parkinson's disease is less than in the general population." This statement applies generally and is contradictory to the actual finding of the study, which is presented in the rest of the item. We should and could have made it clearer that the first statement was intended, as in most Minerva items, as background.